DSC 231 Project Report: Sound-based Bottle Categorizations

Shuying Li
Halıcıoğlu Data Science Institute
La Jolla, CA 92093
shl146@ucsd.edu

Introduction

Recycling is a critical component of waste management, with significant environmental and economic benefits. In the United States, recycling efforts face considerable challenges, contributing to less effective outcomes. Despite 94% of Americans supporting recycling and 74% believing it should be a priority, the national recycling rate remains at only 34.7% [1]. Many Americans struggle with a lack of knowledge about what can be recycled, resulting in contamination of recycling streams and ineffective recycling efforts [2, 3]. Additionally, systemic issues such as inadequate infrastructure and economic barriers further complicate recycling processes [4]. These challenges highlight the importance of improving recycling efficiency to enhance environmental sustainability.

Project Objectives

In consideration of the need to improve recycling processes, the objective of this project is to develop a prototype s shown in the Figure 1 that classifies bottles into three categories—glass, metal, and plastic—before they are processed a. One of the related works on a similar topic was developed by Samuel Alexander [5]. He successfully classified plastic bottles, cans, ping pong balls, and papers by analyzing the sound of their collisions and sorting them into separate internal compartments. Inspired by his project, I decided to focus specifically on bottles.

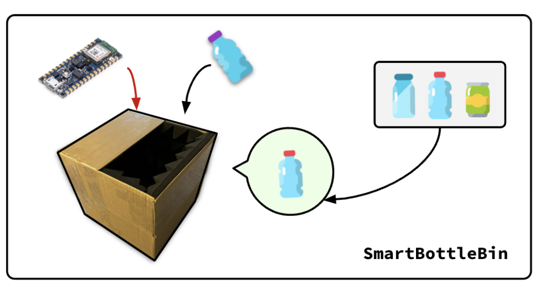

Figure 1: The objective of the prototype

By accurately sorting these different material containers, the prototype aims to enhance the efficiency of recycling operations. The ultimate goal is to reduce contamination in recycling streams, streamline recycling processes, and contribute to more effective waste management practices. Additionally, the system is designed to operate in real-time with low latency, ensuring quick and efficient sorting of materials. This approach not only improves the accuracy of recycling but also minimizes delays in the recycling process, making it more effective.

Project Space & Choice

In this project, the primary focus is on capturing audio and analyzing it in real-time. To achieve this, I utilized the Arduino Learning Kit - TinyML.

Microphone: PDM Microphone in Arduino Nano 33 BLE Sense. It has the ability to capture high-quality audio at a specified sample rate in real time.

Acoustic Environment: We used a personalized collections box with acoustic foam to collect data and make sure there is no external noise. Detail of the collection box will be introduced in the latter section.

Machine Learning Model: This project integrates the Edge Impulse inferencing library to utilize a pre-trained machine learning model for classifying the audio data. The project workflow in Edge Impulse is shown in Figure 2.

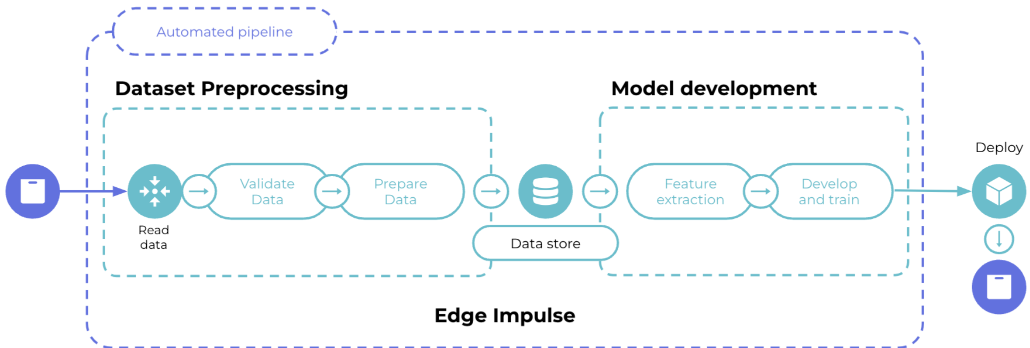

Figure 2: Project workflow by using the Edge Impulse platform

Originally, I planned to directly use TinyML to collect data. However, I found that this approach was not effective and was time-consuming. By using the Edge Impulse platform, data collection became much more user-friendly. More details will be introduced in the latter section.

The Edge Impulse project for this project is the following link: dsc231-final-project.

Data Collections

The data collection in this project is the most important parts. The quality of audio data directly affect the accuracy of model. In this section, I will introduce the details of the collection box and provide details about the types of bottles used in this project.

Bottle Types

Figure 3: Bottle type illustrations

In this project, I selected different bottles for each category—glass, plastic, and metal or can. The reason for choosing various bottles within the same category is to account for differences in size, height, weight, and material hardness, which can influence the collision sounds even within the same category. The differences in physical properties among the bottles within the same category will help to determine the robustness of the audio classification model.

During the data collection process, samples were recorded with a length of 10,000 milliseconds at a frequency of 16,000 Hz. Each collision sound was captured in segments of 1,000 milliseconds using the built-in microphone sensor from the BLE 33 Sensor. All segmentation was finished in Edge Impulse.

Collection Box

The purposes of creating the collection box are:

- To provide a controlled environment for recording hitting sounds without interference from external noises.
- To simulate the conditions of a real-life rubbish bin.

The collection box as shown in the Figure 4, or prototype for SmartBottleBin, is constructed from a cardboard box measuring 12" x 12". The interior of the box is lined with acoustic foam on all four sides to minimize external noise. At the bottom, there is an impact plate or the hitting surface made of high-density polyethylene (HDPE) that serves as the hitting surface for the bottles.

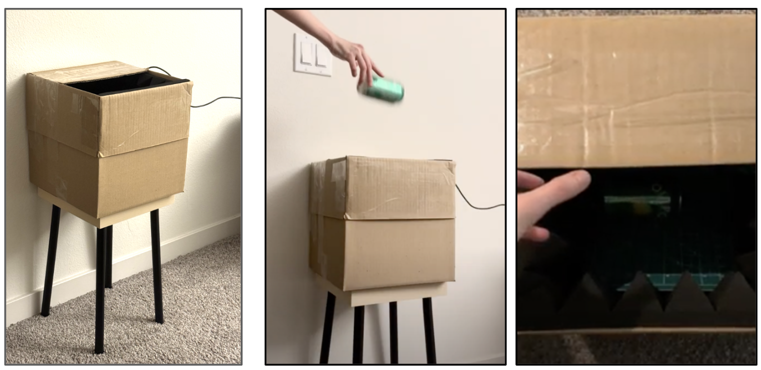

Figure 4: The first two figures show the overall appearance of the collection box, including the opening used for discarding bottles. The BLE 33 Sensor is located inside the box and is connected to the laptop. The third figure provides an inside view of the collection box, showing the acoustic foam lining and the impact plate.

Impact Plates

Impact plates or the grounding surfaces that reflect collision sounds, also play a important role in this project. The material composition and positioning of these impact plates can influence the resonance and vibrations produced during collisions, thereby affecting the experimental results.

Materials Choosing an appropriate material can make the sound of bottles made from three different materials more distinctive. During the initial phase of this project, I experimented with various materials: polypropylene plastic (PP), like an IKEA rubbish bin, foam board, polyvinyl chloride (PVC), and high-density polyethylene (HDPE), similar to the material used in cutting boards. During my experiment, I positioned each material plate horizontally and observed the sounds produced when bottles made from different materials hit the plate. The criteria I used to determine which material is better are based on: 1) The ability to clearly hear the differences in the sounds produced by human and 2) The Mel Frequency Cepstral (MFE) analysis of the sounds. During my experiments, I found that HDPE is the best material for distinguishing between the three different types of bottles. As shown in the Figure 5, the comparison of each material demonstrates that HDPE provides the clearest and most distinct sound profiles for cans, plastic, and glass bottles.

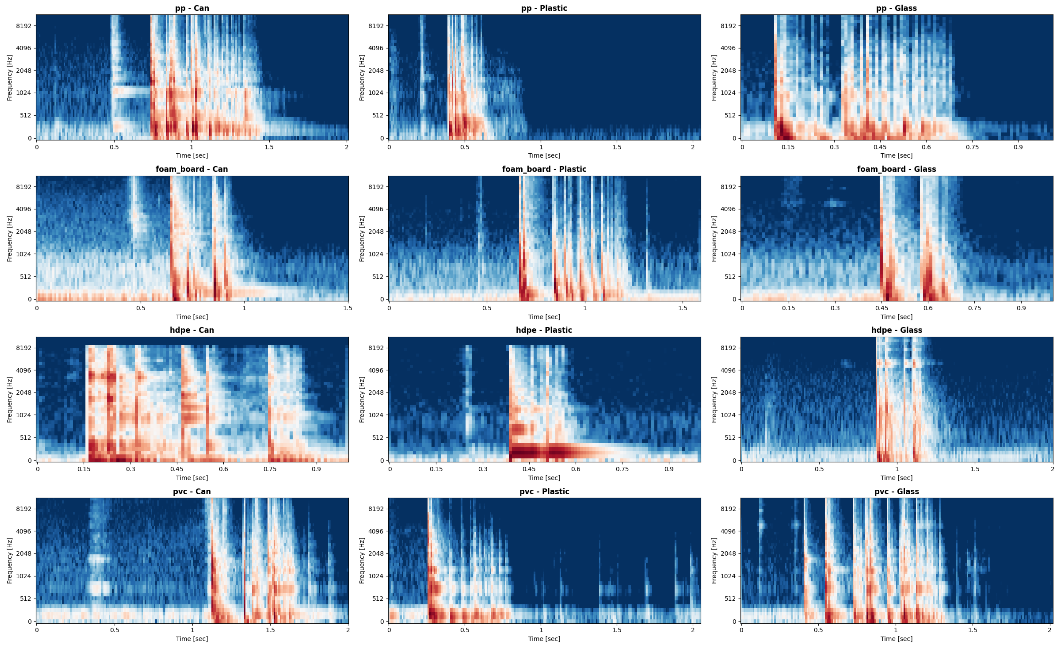

Figure 5: Comparison of each materials

Impact Plates Positions One of the limitations of the collection box is its small size, especially after acoustic foam is added. This restricted space means that using a horizontal position for the impact plates does not accurately mimic real-world conditions because most street rubbish bin have large volume and relative tall heights. To address this, I employed different positions for the impact plates as shown in the Figure 6 to better simulate the actual environment. By varying the angles and orientations of the impact plates, the collection box can more effectively capture and analyze the sounds of bottles hitting the surface.

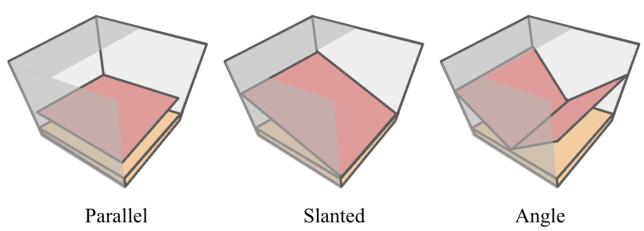

Figure 6: Different positions of impact plate

Throwing Methods

The manner in which people throw bottles can produce different collision sounds. To accurately simulate real-world conditions, I utilized four throwing methods in this project as shown in Figure 7.

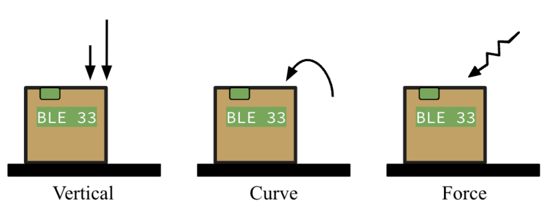

Figure 7: Throwing method illustration

The first two methods are throwing the bottle vertically from different heights. For the curve, it means that the bottle is thrown with a curved trajectory. The last method is thrown with additional force, simulating a more aggressive or forceful disposal.

Experiments

Dataset Insights

In this project, I collected data as follows: cans (43 samples), glass (42 samples), noise (49 samples), and plastic (40 samples). 80% of the collected data was used for training the model, while the remaining 20% was used for testing.

For audio preprocessing and feature extraction, the project utilized the Audio MFE (Mel Frequency Cepstral) processing block. This method extracts time and frequency features from the audio signal, similar to a spectrogram, but employs a non-linear scale in the frequency domain known as the Mel-scale. This method is great for non-voice audio [6].

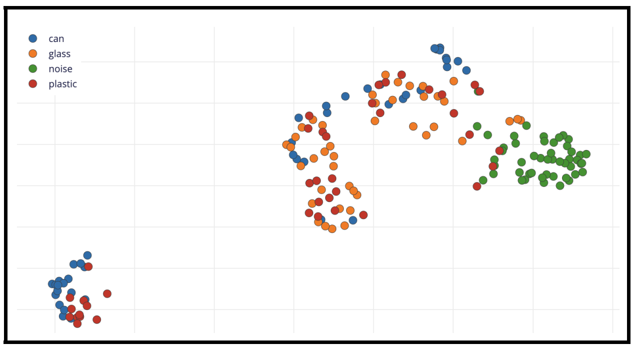

Figure 8: The figure visualizes the clustering of different bottle types—can, glass, noise, and plastic—based on their audio features extracted using Mel Frequency Cepstral Coefficients (MFCC). This figure is generated in Edge Impulse platform.

After using MFE to extract features from the audio samples, as shown in Figure 8, it shows that certain plastic, can, and almost all noise samples have distinctive characteristics. However, there are some overlaps between glass, can, and plastic samples. This overlap may be due to several factors:

- Throwing Method: Different throwing methods can affect the sound produced by the bottles, leading to variations in the audio features.
- Material: Hard plastic bottles may have an acoustical fingerprint similar to that of glass, causing overlap in their audio characteristics.

Model Deployments

The model deployment was developed using Edge Impulse. Data augmentations were not included in this project. One important factor to consider is the model size, given that our device has a limited RAM size of 340 KB and ROM is 1024 KB, therefore complicated model is not the optimal solutions.

The baseline model is a 1D Convolutional Neural Network with the following architecture:

- Input layer: 3,960 features
- Reshape layer: 40 columns
- 1D Convolutional/Pooling layer: 8 neurons, kernel size of 3, 1 layer
- Dropout layer: Rate of 0.25
- 1D Convolutional/Pooling layer: 16 neurons, kernel size of 3, 1 layer
- Dropout layer: Rate of 0.25

- Flatten layer
- Add an extra layer
- Output layer: 4 classes

The EON Tuner is also applied to optimize and select the best machine learning model for embedded applications. It significantly lowers the barrier to entry for advanced model generation and optimization, making it easier for developers to deploy efficient machine learning models on edge devices [7].

Deployments in Arduino Device: After developing the model, it will be converted to an Arduino library that contains model.

Results

After model developments, the result is shown in the Table 1.

Model	Accuracy	Loss	epoch	RAM	ROM	Latency
1D convolutional	94.3%	0.12	100	14 kB	35 kB	34 ms
1D convolutional	100.0%	0.01	500	16 kB	44 kB	24 ms
2D convolutional	89%	0.255	100	17 kB	34 kB	25 ms

Table 1: Model comparisons

The 1D convolutional model trained for 500 epochs provides the best overall performance in terms of accuracy and loss, but at the cost of slightly higher RAM and ROM usage. It also achieves the lowest latency, making it suitable for real-time applications. The 1D convolutional model trained for 100 epochs, while less accurate, is the most resource-efficient, making it ideal for deployment on devices with strict memory constraints. The 2D convolutional model, although having decent performance, does not match the accuracy of the 1D models and uses more RAM. It might be less preferable unless the application specifically benefits from 2D convolutions. Therefore, the final model is the 1D convolutional model trained for 100 epochs.

Deployments in Arduino Device: After deploying the model on the kit, the inference time is real-time, and recording begins 2 seconds after the inference is completed.

Limitations

Data Aspects In this project, I only collected data for three types of bottles made from different materials. However, in the real world, there is a greater variety of container types that should be considered to improve the robustness of the model.

Future work should include:

- Additional Material Types: Incorporate Tetra Pak as a fourth class to cover a broader range of common recyclable materials.
- Environmental Noise: Add human voices and other ambient sounds to the dataset to simulate more realistic environmental conditions.
- Diverse Data Collection: Increase the variety and quantity of data by collecting samples from different types of bottles, including variations in size, shape, and material to enhance the model's generalization capabilities.

Data Collections The collection box used in this project is relatively small. For future iterations, it is important to use a larger box to better simulate real-world conditions and ensure the results are more representative of typical usage scenarios.

References

[1] PIRG. It's "america recycles day." is the united states set up for recycling success? 2018.

[2] Waste Advantage Magazine. Why americans aren't recycling. 2024.

[3] TechXplore. Report finds americans throw three-quarters of their recyclables into the trash. 2024.

[4] State of the Planet. Recycling in the u.s. is broken. how do we fix it? 2024.

[5] Arduino Team. This recycling bin sorts waste using audio classification, 2023. Accessed: 2024-06-11.

[6] Buket D. Barkana, Burak Uzkent, and Inci Saricicek. Normal and abnormal non-speech audio event detection using mfcc and pr-based feature sets. *Advanced Materials Research*, 601:200–208, 2012.

[7] Edge Impulse. Introducing the eon tuner: Edge impulse's new automl tool for embedded machine learning, 2021. Accessed: 2024-06-12.